Machine Learning for Anatomical Neuroimaging

John Ashburner

Wellcome Trust Centre for Neuroimaging,
UCL Institute of Neurology,
12 Queen Square,
London WC1N 3BG,
UK.

To explain or to predict?

"*The only relevant test of the validity of a hypothesis is comparison of prediction with experience.*"

Milton Friedman

To explain or to predict?

"The predictive point of view is a prototypical point of view to explain the basic activity of statistical analysis."

— Akaike

"The only useful function of a statistician is to make predictions."

— Deming

"The prediction of observables or potential observables is of much greater relevance than the estimate of what are often artificial constructs-parameters"

— Geisser

Galit Shmueli. "To Explain or to Predict?". Statistical Science 25(3):289–310 (2010).

To explain or to predict?

You can interpret this in the sense that the **construct or inventions that you make must be of such a kind that the consequences that you compute are comparable with experiment** *- that is, that you do not compute a consequence like a 'moo must be three goos', when nobody knows what a moo or a goo is. ...*
*....***Science is only useful if it tells you about some experiment that has not been done; it is no good if it only tells you what just went on.**

— Richard Feynman; The Character of Physical Law

Evidence-based Science

...also just known as "science".

- Researchers claim to find differences between groups. Do those findings actually discriminate?
- How can we most accurately diagnose a disorder from image data?
- Pharma wants biomarkers. How do we most effectively identify them?
- There are lots of potential imaging biomarkers. Which are most (cost) effective?

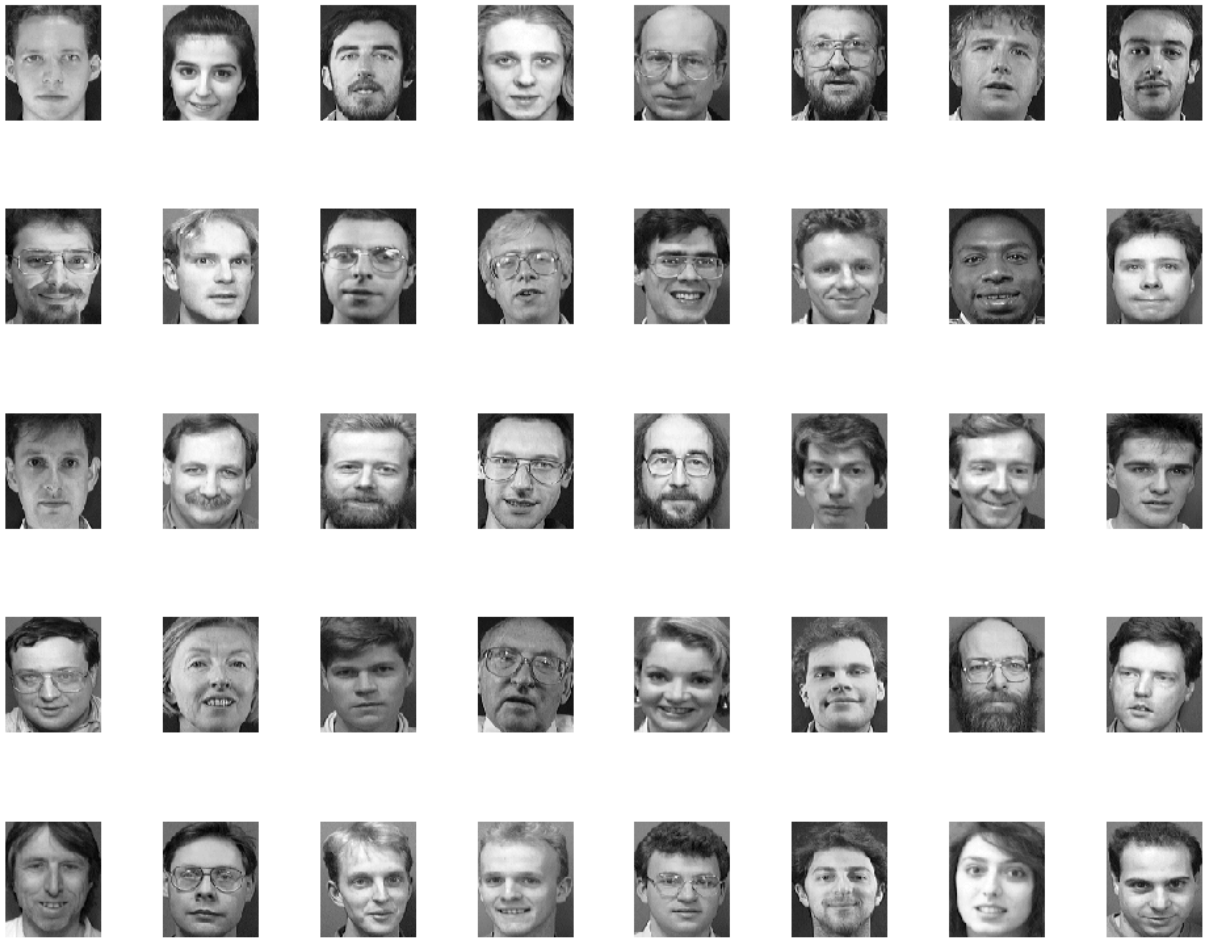
Pattern recognition provides a framework to compare data (or preprocessing strategy) to determine the most accurate approach.

Inter-subject Variability

Why focus on anatomy?

- Many medical applications involve understanding differences among individuals/populations.
- In image data, most of the differences we can see are anatomical in nature.
- Understanding growth and development requires us to look at growth and development (anatomy).

Biological variability is multivariate



Cause and Effect

- Decisions about ....

Decision-making under uncertainty

- Clinical decision-making is like gambling.
- Optimal way to gamble is Bayesian.
- "Dutch Book Theorem".

No Free Ducklings

No Free Lunch theorem says that learning is impossible without prior knowledge (`http://en.wikipedia.org/wiki/No_free_lunch_in_search_and_optimization`).
Ugly Duckling theorem says that things are all equivalently similar to each other without prior knowledge (`http://en.wikipedia.org/wiki/Ugly_duckling_theorem`).

What prior knowledge do we have about the variability among people that can be measured using MRI?
How do we use this knowledge?

Different ways of measuring distances

David Mumford

Mathematician

David Bryant Mumford is an American mathematician known for distinguished work in algebraic geometry, and then for research into vision and pattern theory. He won the Fields Medal and was a MacArthur Fellow. Wikipedia

Born: June 11, 1937 (age 76), Worth village, West Sussex, Crawley

Children: Steve Mumford

Education: Phillips Exeter Academy, Harvard University

Awards: Fields Medal, Wolf Prize in Mathematics, MacArthur Fellowship, The Shaw Prize in Mathematical Sciences, National Medal of Science for Mathematics and Computer Science

Empirical Statistics and Stochastic Models for Visual Signals

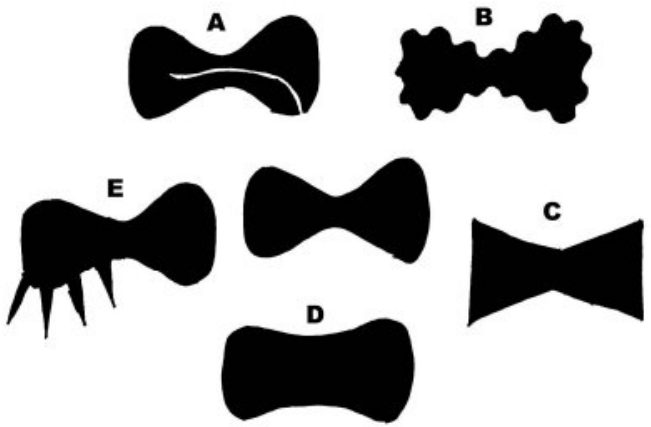


Figure 1.11 Each of the shapes A,B,C,D and E is similar to the central shape, but *in different ways*. Different metrics on the space of shape bring out these distinctions.

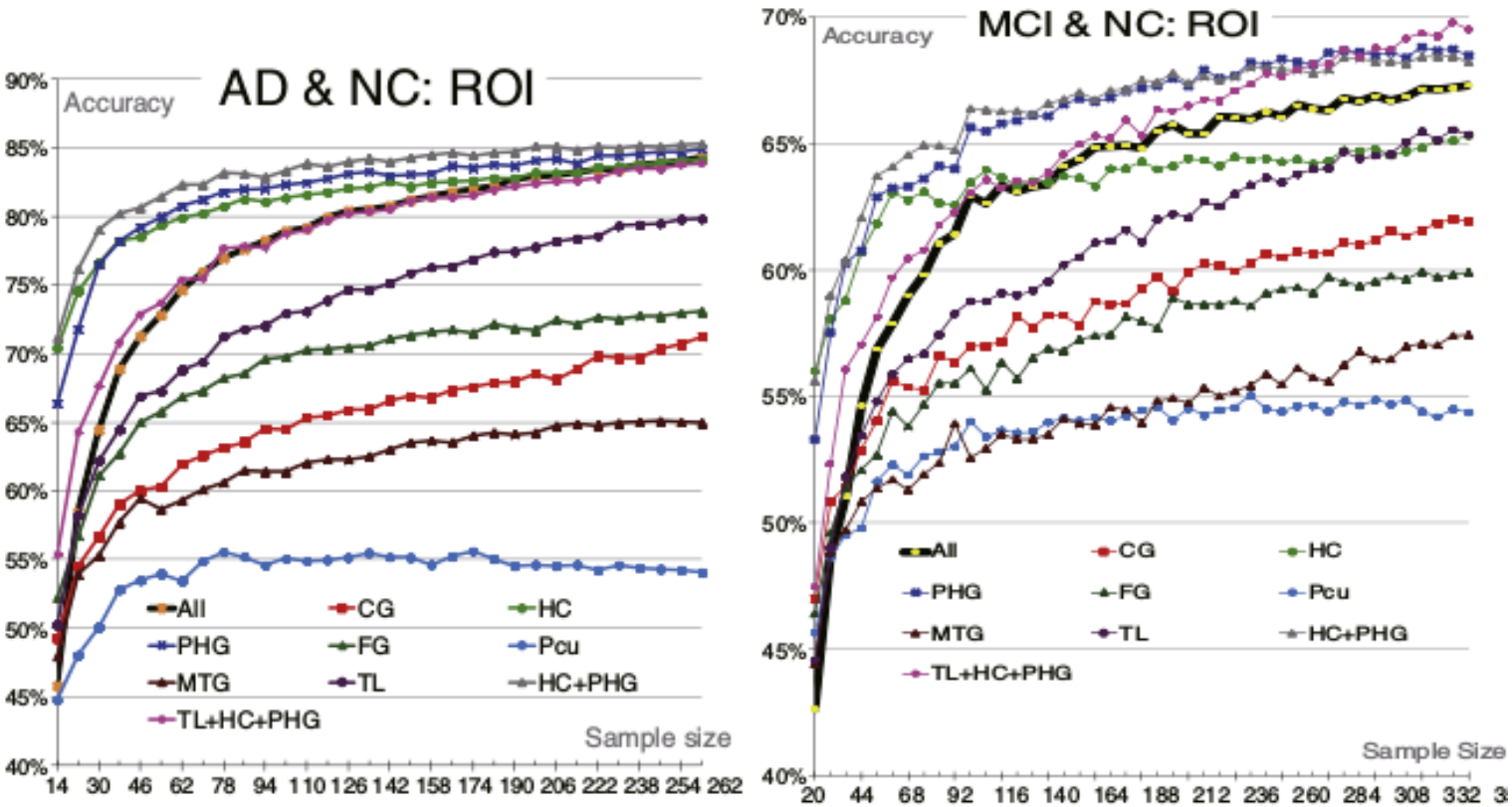
Weighting the data

$$\mathbf{K} = \mathbf{X}\mathbf{W}\mathbf{X}^T$$

Prior knowledge about brain regions involved

- The best way would be to augment the training data with data from previous studies.
- Lack of data-sharing means this is generally not possible, so we need to extract information from publications.
- The neuroimaging literature is mostly blobs.
- These give pointers about how best to weight the data ($\mathbf{W} = diag(\mathbf{w}), w_i \in \mathbb{R}^+$).

Weighting suspected regions more heavily

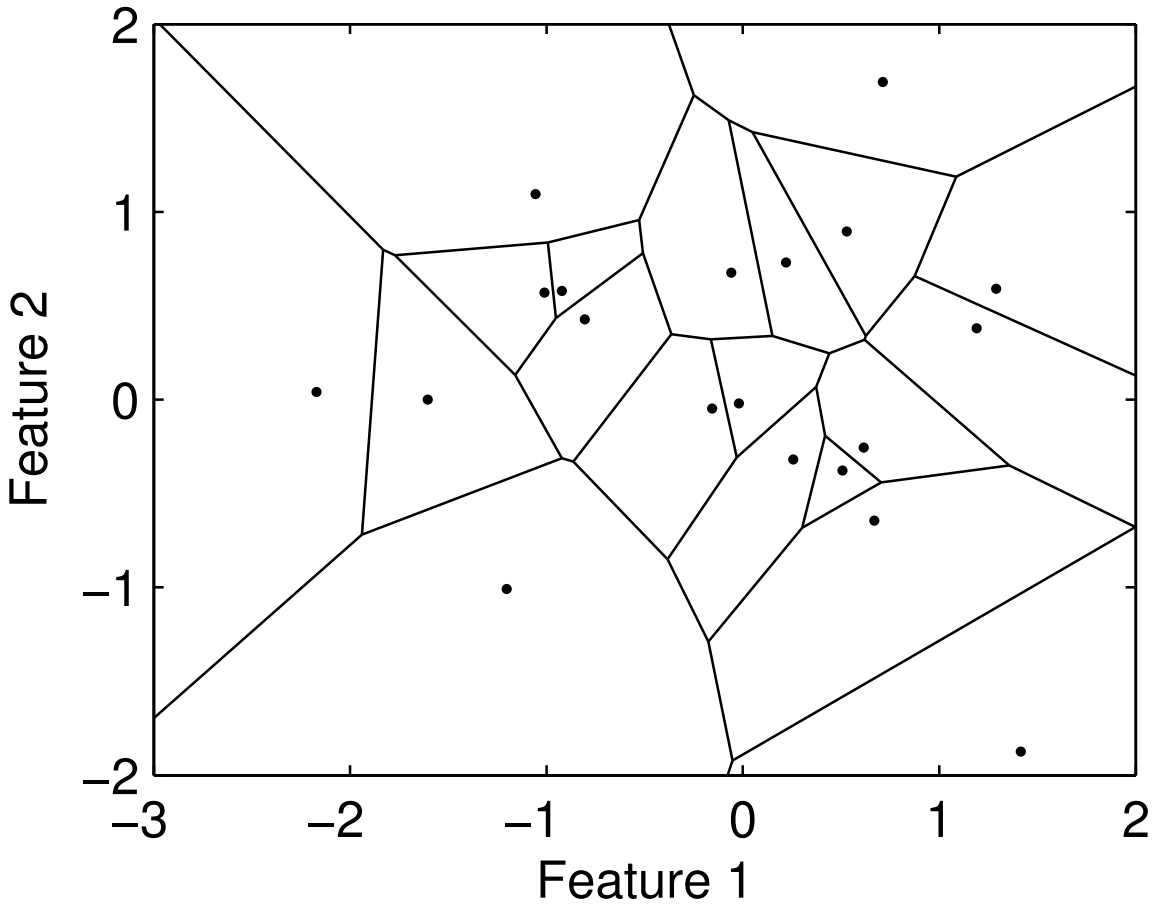


Chu et al. "Does feature selection improve classification accuracy? Impact of sample size and feature selection on classification using anatomical magnetic resonance images". NeuroImage 60:59–70 (2012).

Curse of dimensionality

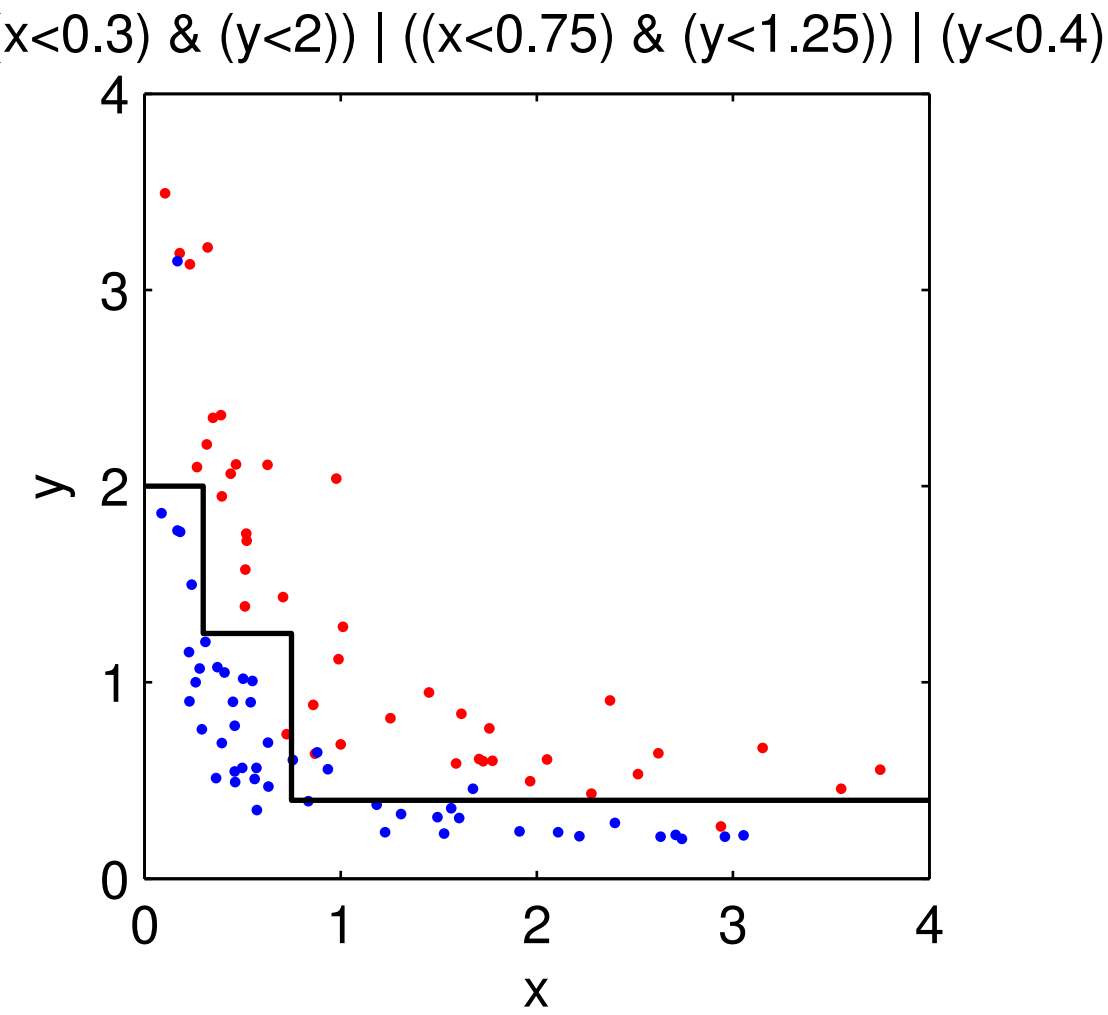
Large p, small n.

Nearest-neighbour classification



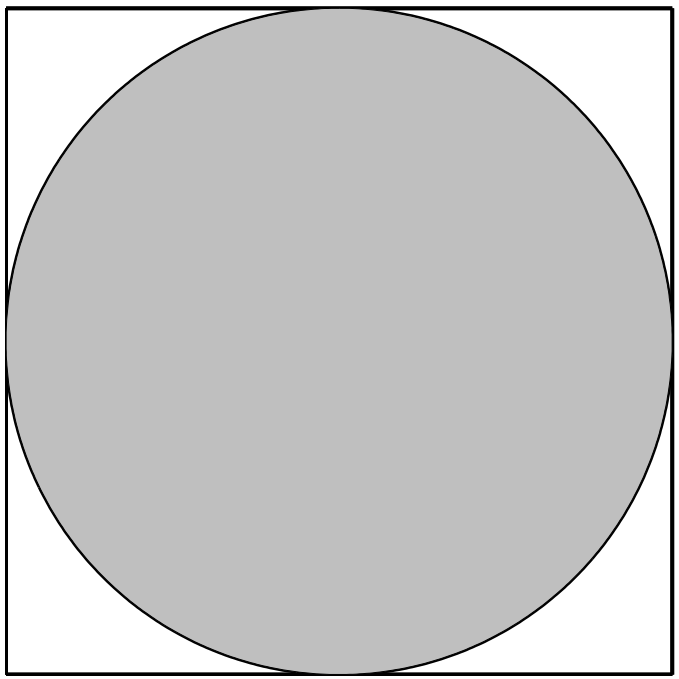
- Not nice smooth separations.
- Lots of sharp corners.
- May be improved with *K-nearest neighbours*.

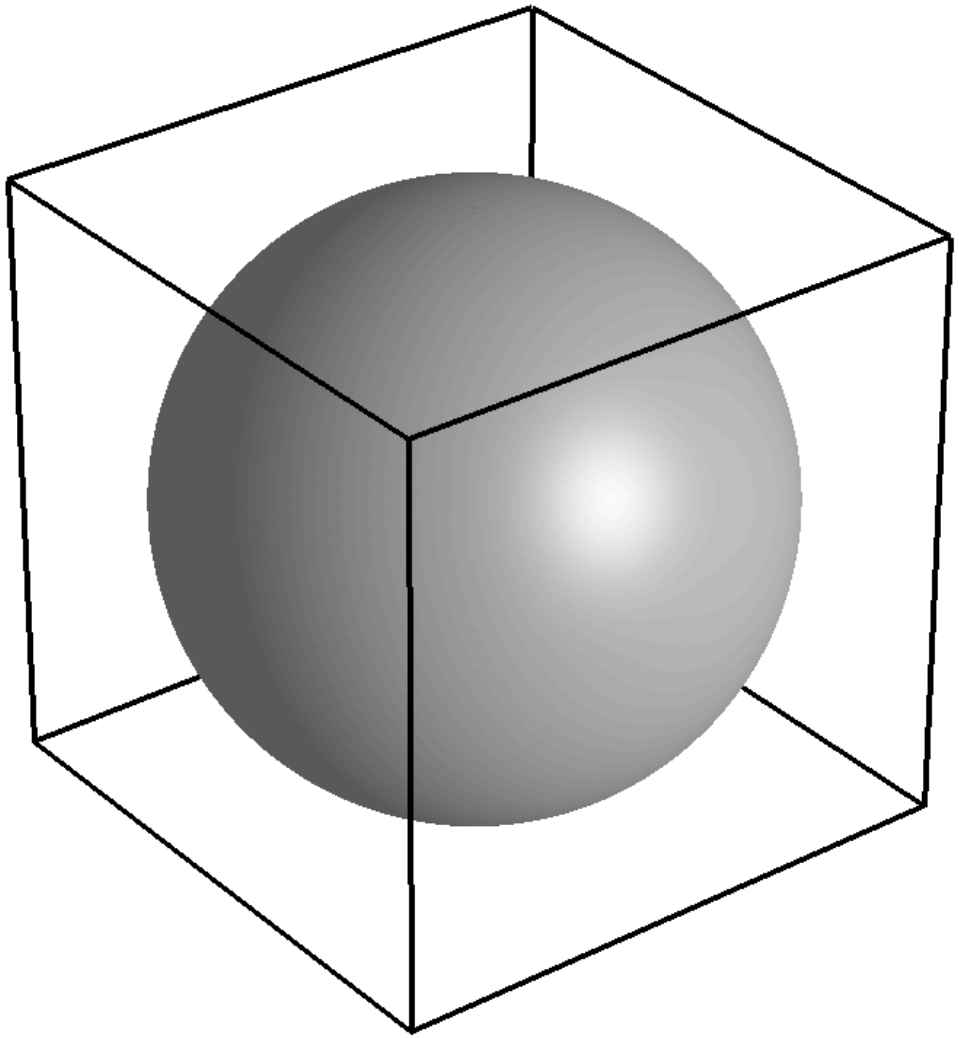
Rule-based approaches



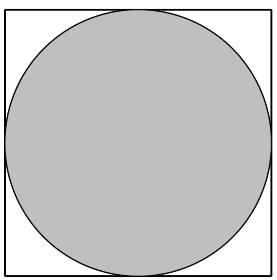
- Not nice smooth separations.
- Lots of sharp corners.

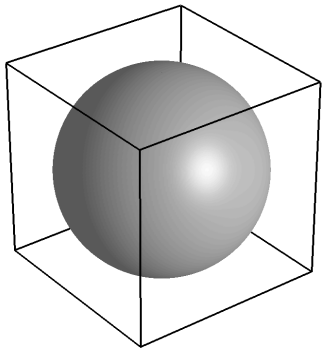
Corners matter in high-dimensions

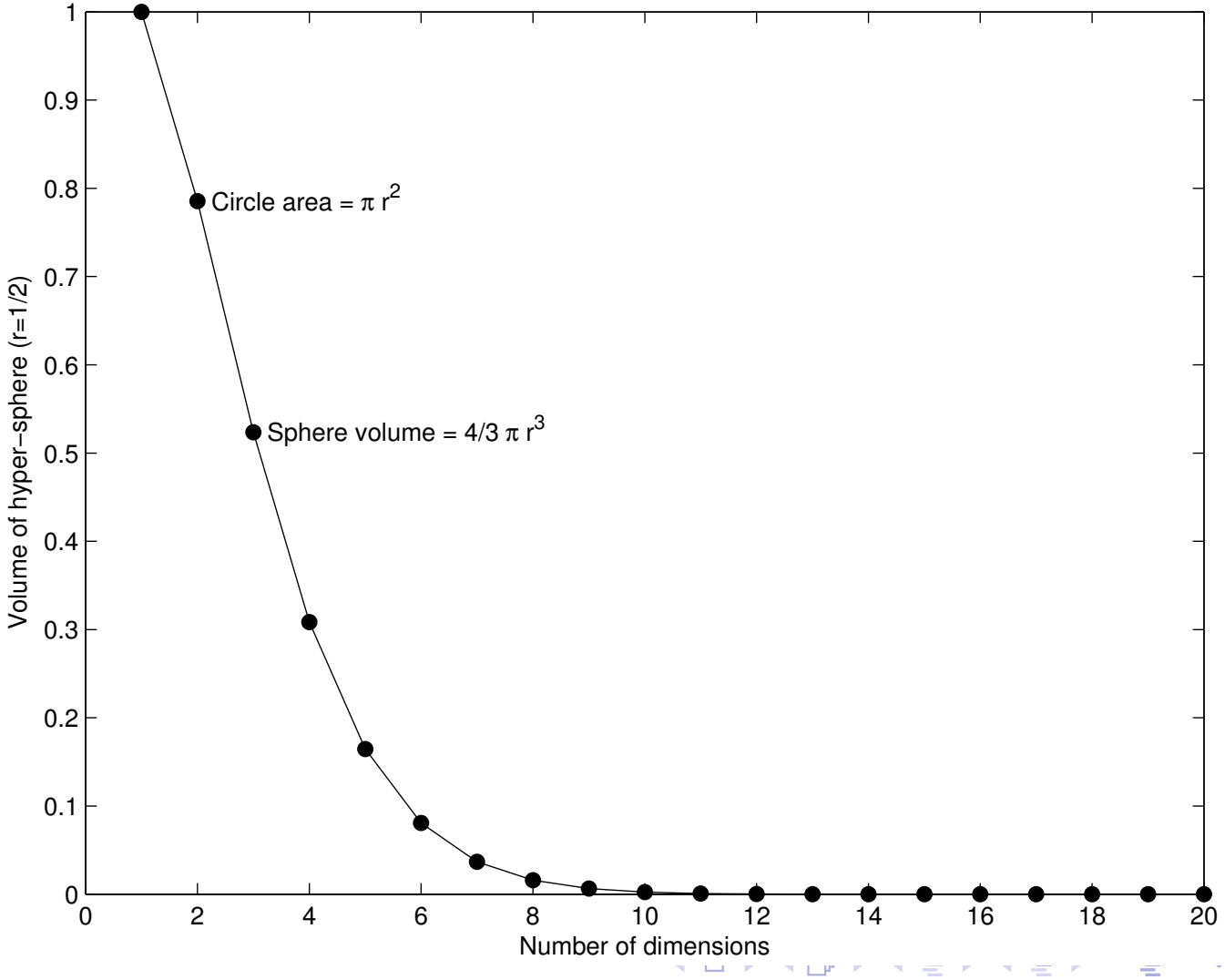




Corners matter in high-dimensions







Dimensionality $\neq$ number of voxels

- Lots of effort on data-driven feature selection methods.
  - Involves estimating weighting matrix such that $\mathbf{W} = diag(\mathbf{w}), w_i \in \{0, 1\}$.
  - Lots of parameters needed to achieve this.
- Many papers claim excellent results.
- Little evidence to suggest that most voxel-based feature selection methods help.
  - Little or no increase in predictive accuracy.
  - Commonly perceived as being more "interpretable".

PITTSBURGH

BRAIN ACTIVITY INTERPRETATION COMPETITION 2007

Competition Home
News & Updates
Competition Overview
Registration
Competition Materials
Literature
Message Board
Contact Information
Download Data
Submit Results

2007 PBAIC Results

First, second, and third place winners were announced at the Competition Workshop at the Organization for Human Brain Mapping being held June 10-14, 2007 in Chicago, Illinois, USA. Prizes will be awarded (1st: $10,000; Tied-2nd: $3,500; 3rd: $2,000; and Neuroscience Award: $5,000). Awards were presented during the morning workshop on Thursday, June 14, 2007. For more information, visit the OHBM website.

Audio | Video | Poster | Slides | Photos

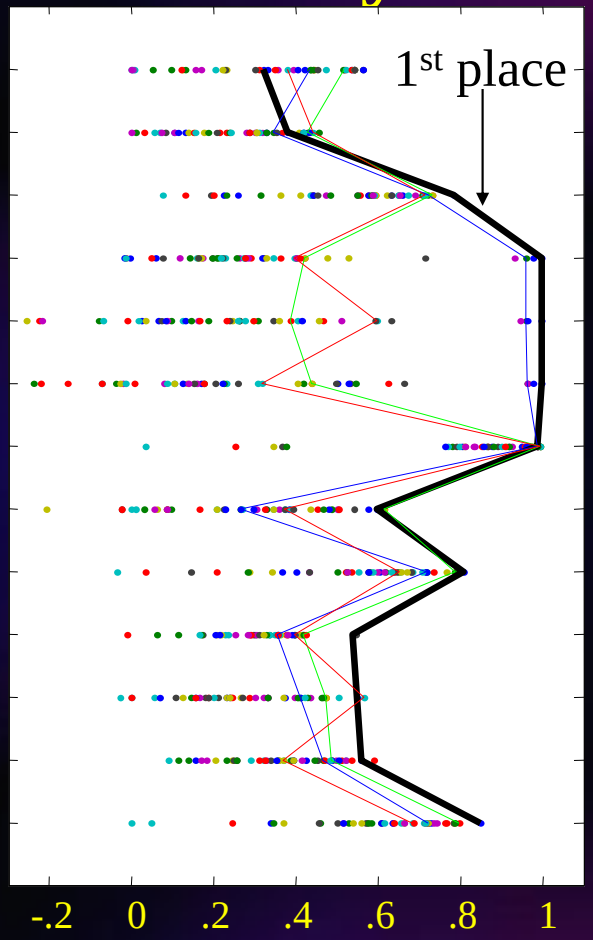
Top 3 Winners

1st Place: $10,000 Prize
Carlton Chu, Yizhao Ni, Geoffrey Tan, John Ashburner
University College London
Title: *Kernel Methods for fMRI Pattern Prediction – applications of Relevance Vector Regression and Kernel Ridge Regression*

2nd Place (tied): $3,500 Prize
Denis Chigirev and The Princeton EBC Team
Princeton University
Title: *One Size Does Not Fit All: Regressor and Subject Specific Techniques for Predicting Behavior in a Structured Environment*

2nd Place (tied): $3,500 Prize
Giancarlo Valente, Federico De Martino, Fabrizio Esposito and the Maastricht Team
University of Maastricht
Title: *Predictions of PBAIC 2007 Ratings with Linear Relevance Vector Machine regression*

There was variability among entrants



Max Correlation over the 3 subjects for each entrant

Life lessons….

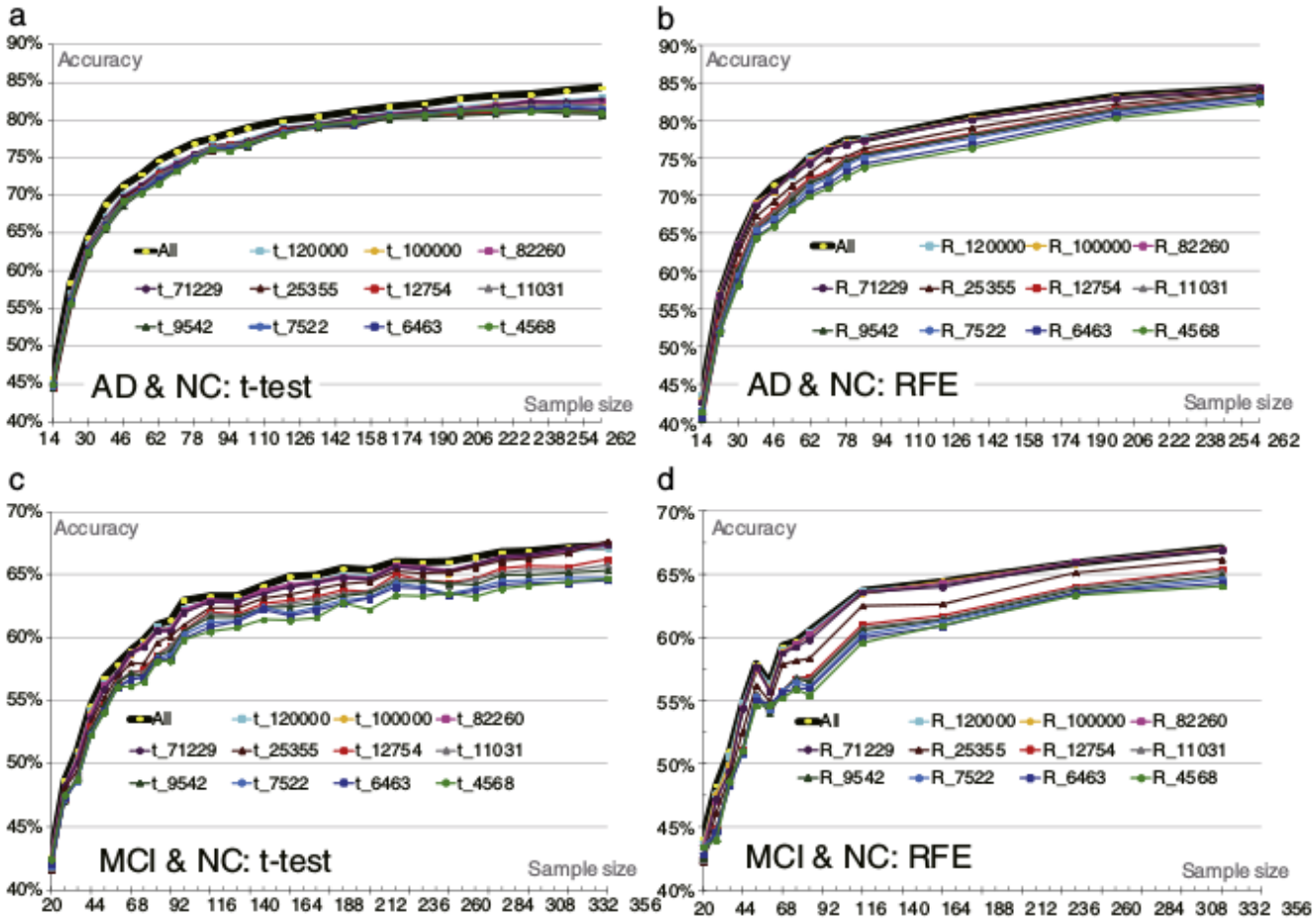
- Very different methods gave similar scores (based on pre- and post-processing)
- Similar methods (e.g., support vector machines) gave very different results.

Data-driven feature selection

"In our evaluation, two methods included a feature selection step: Voxel-STAND and Voxel-COMPARE. Overall, these methods did not perform substantially better than simpler ones... ... A more robust way to decrease the dimensionality of the features way would be to use more prior knowledge of the disease."

Cuingnet et al. "Automatic classification of patients with Alzheimer's disease from structural MRI: A comparison of ten methods using the ADNI database". NeuroImage 56(2):766–781 (2011).

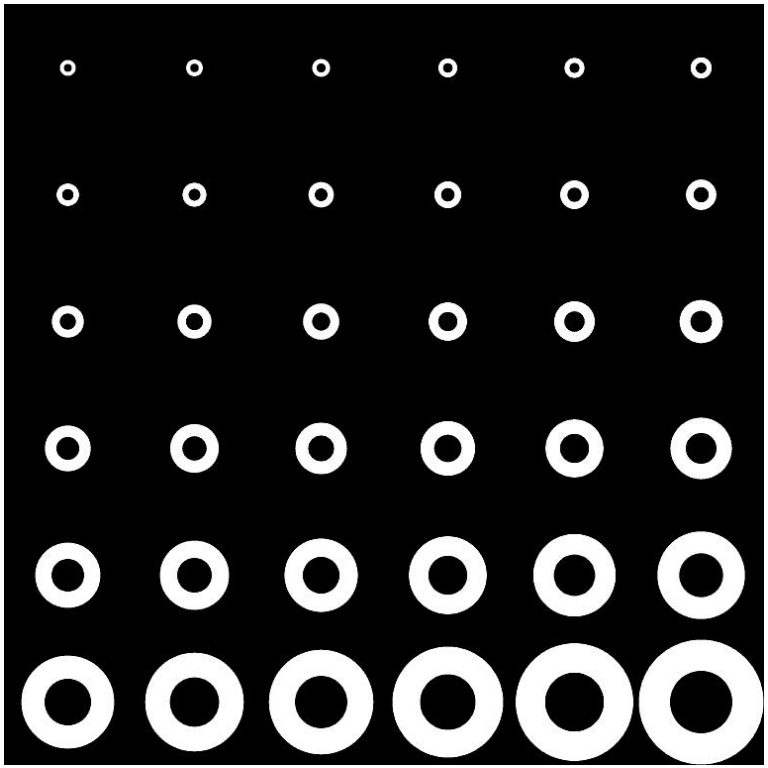
Data-driven feature selection



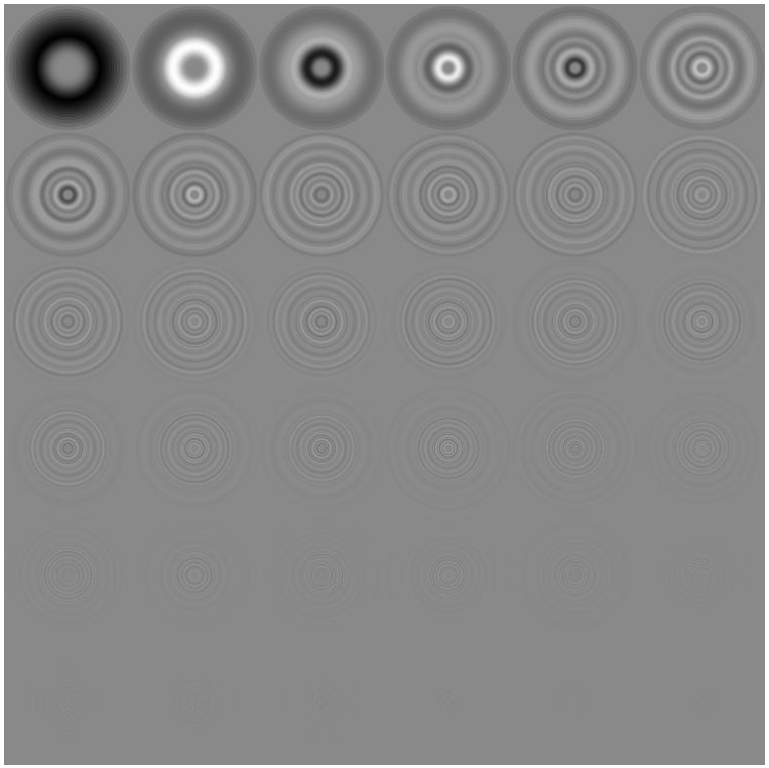
Chu et al. "Does feature selection improve classification accuracy? Impact of sample size and feature selection on classification using anatomical magnetic resonance images". NeuroImage 60:59–70 (2012).

One mode of geometric variability

Simulated images



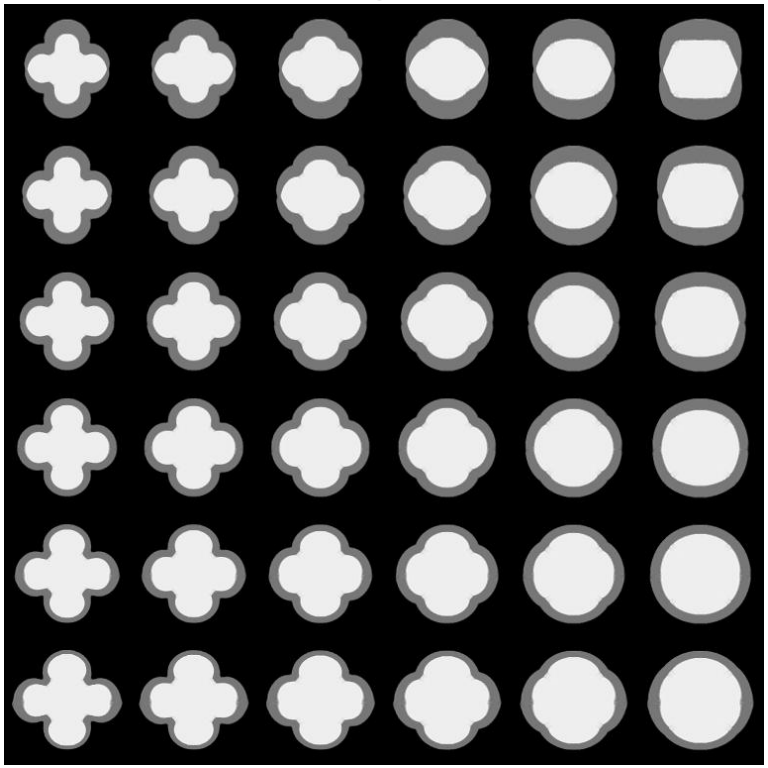
Principal components



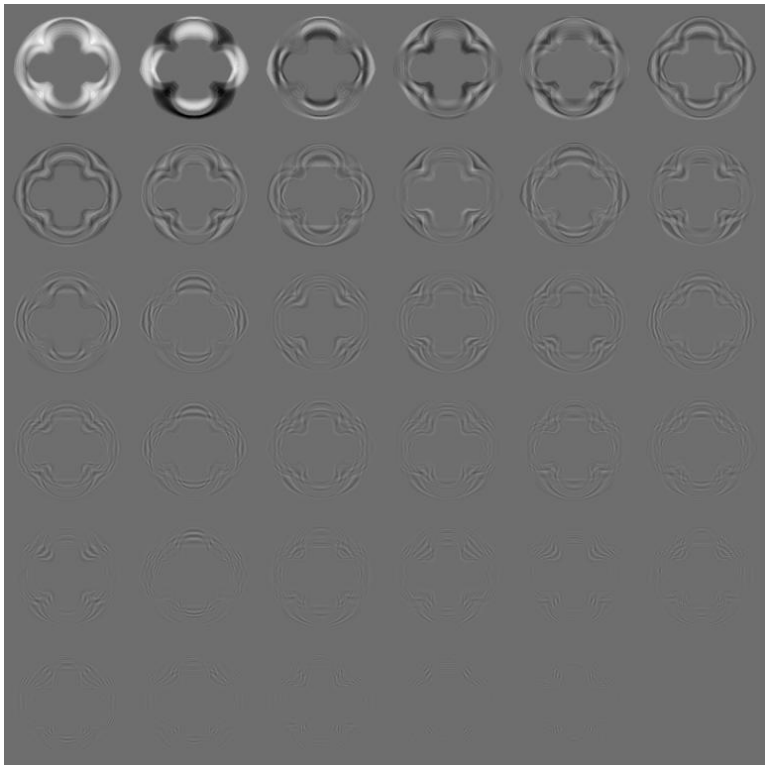
A suitable model would reduce these data to a single dimension.

Two modes of geometric variability

Simulated images



Principal components

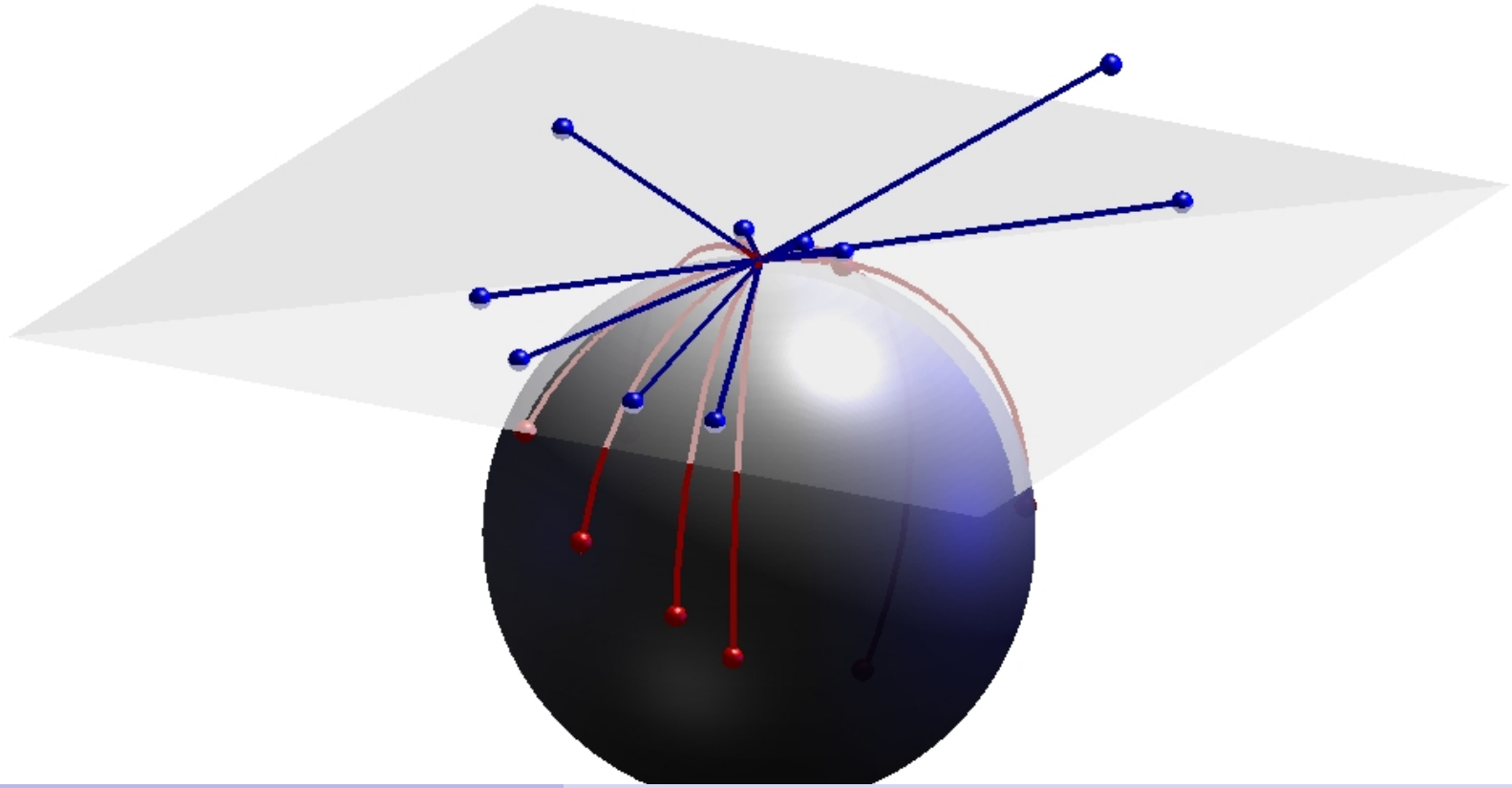


A suitable model would reduce these data to two dimensions.

Non-Euclidean geometry

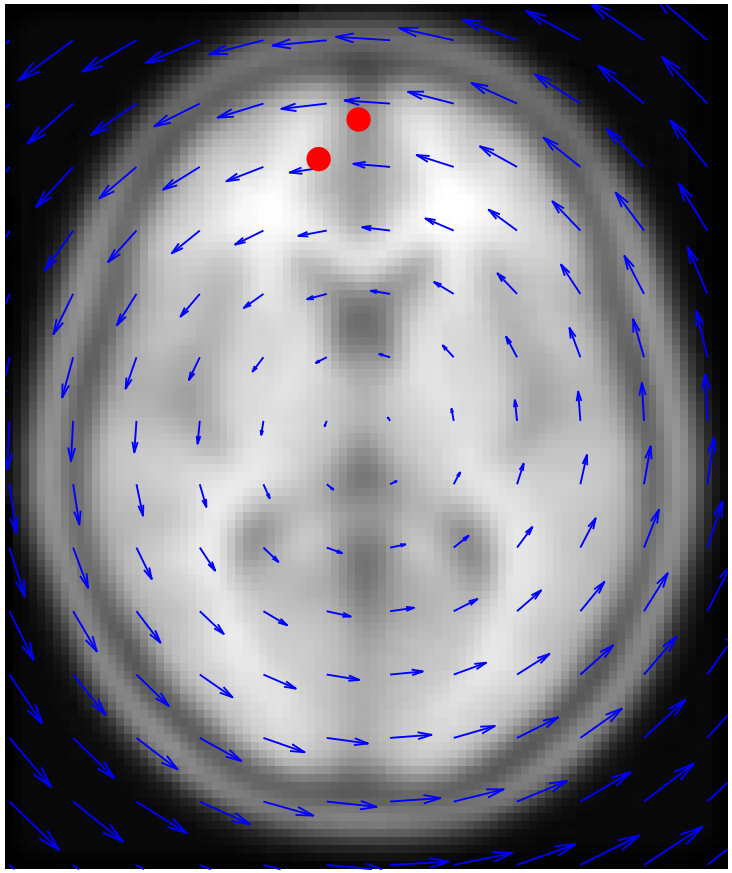
- Distances are not always measured along a straight line.
- "*Shapes are the ultimate non-linear sort of thing*"

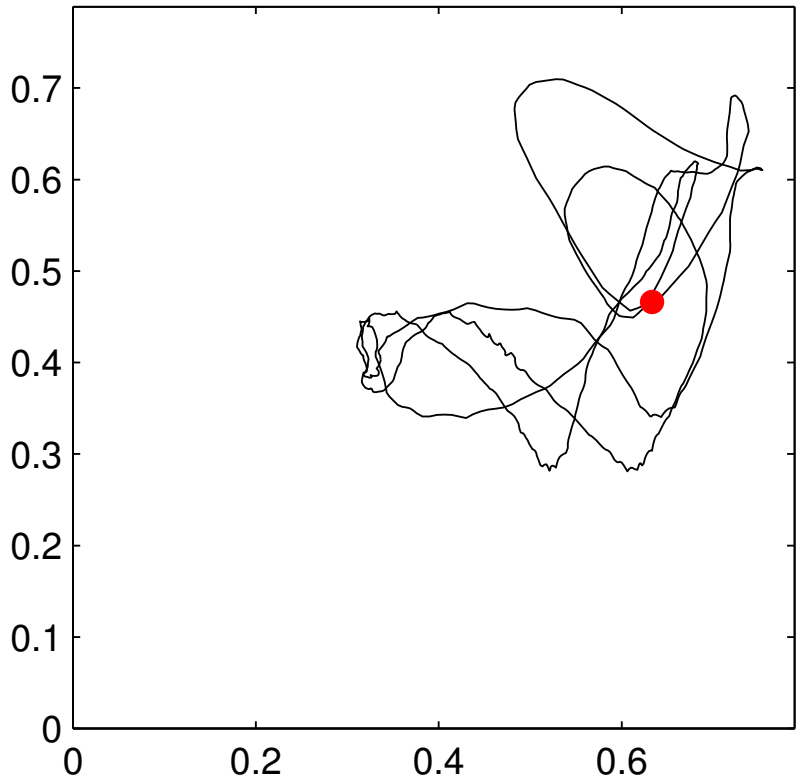
Linear approximations to nonlinear problems



Transformed images fall on manifolds

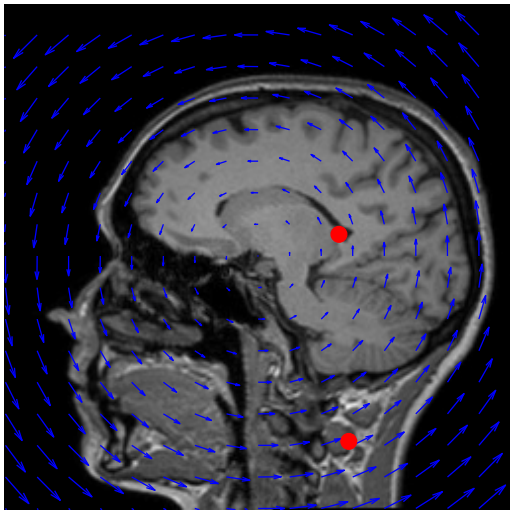
Rotating an image leads to points on a 1D manifold.

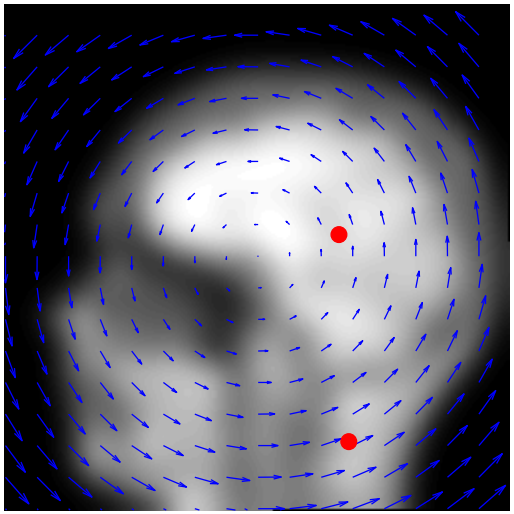


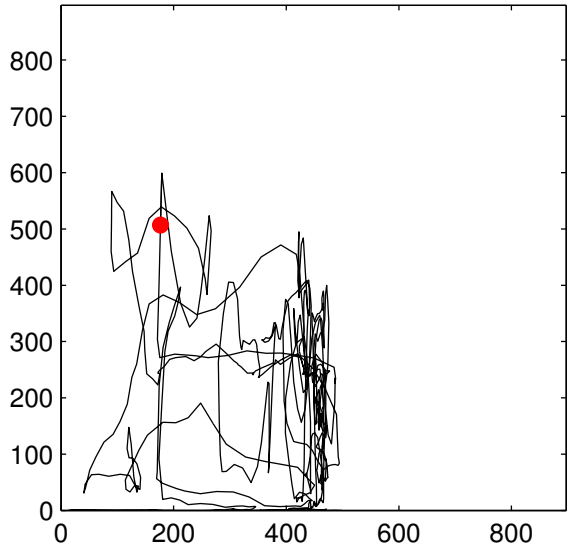


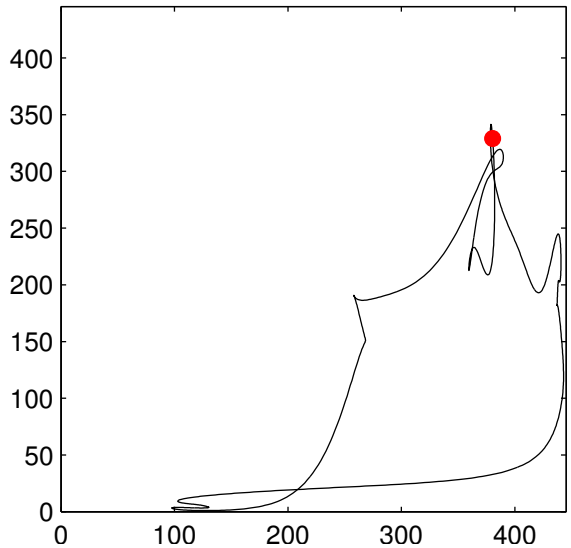
Rigid-body motion leads to a 6-dimensional manifold (not shown).

Local linearisation through smoothing









Spatial smoothing can make the manifolds more linear with respect to small misregistrations. Some information is inevitably lost.

Metrics

Distances need to satisfy the properties of a *metric*:

1. $d(\mathbf{x}, \mathbf{y}) \geq 0$ (non-negativity)
2. $d(\mathbf{x}, \mathbf{y}) = 0$ if and only if $\mathbf{x} = \mathbf{y}$ (identity of indiscernibles)
3. $d(\mathbf{x}, \mathbf{y}) = d(\mathbf{y}, \mathbf{x})$ (symmetry)
4. $d(\mathbf{x}, \mathbf{z}) \leq d(\mathbf{x}, \mathbf{y}) + d(\mathbf{y}, \mathbf{z})$ (triangle inequality).

Satisfying (3) requires inverse-consistent image registration.
Satisfying (4) requires a specific class of image registration models.

Non-Euclidean geometry

- Distances are not always measured along a straight line.
- "*Shapes are the ultimate non-linear sort of thing*"

...all clinically diagnosed AD patients will not have AD pathology, and up to 30% of cognitively normal subjects will have a pathologic diagnosis of AD at autopsy.

Vemuri et al. "Antemortem MRI based STructural Abnormality iNDex (STAND)-scores correlate with postmortem Braak neurofibrillary tangle stage". NeuroImage 42(2):559–567 (2008).